

8 *Greenberg's Indo-Pacific hypothesis: an assessment*

ANDREW PAWLEY

1 Introduction

In 1971 Joseph Greenberg published evidence for his Indo-Pacific hypothesis, which proposed that there is a genetic relationship between all the non-Austronesian language families of Melanesia and the Halmahera and Timor regions of the Indonesian archipelago, together with the languages of Tasmania and most if not all the languages of the Andaman Islands.¹ The main evidence consisted of 84 sets of resemblant words ('Indo-Pacific etymologies') plus some resemblances in grammatical elements, almost every set being represented in at least three of the 14 groups which he treated, provisionally, as primary branches of Indo-Pacific (IP).

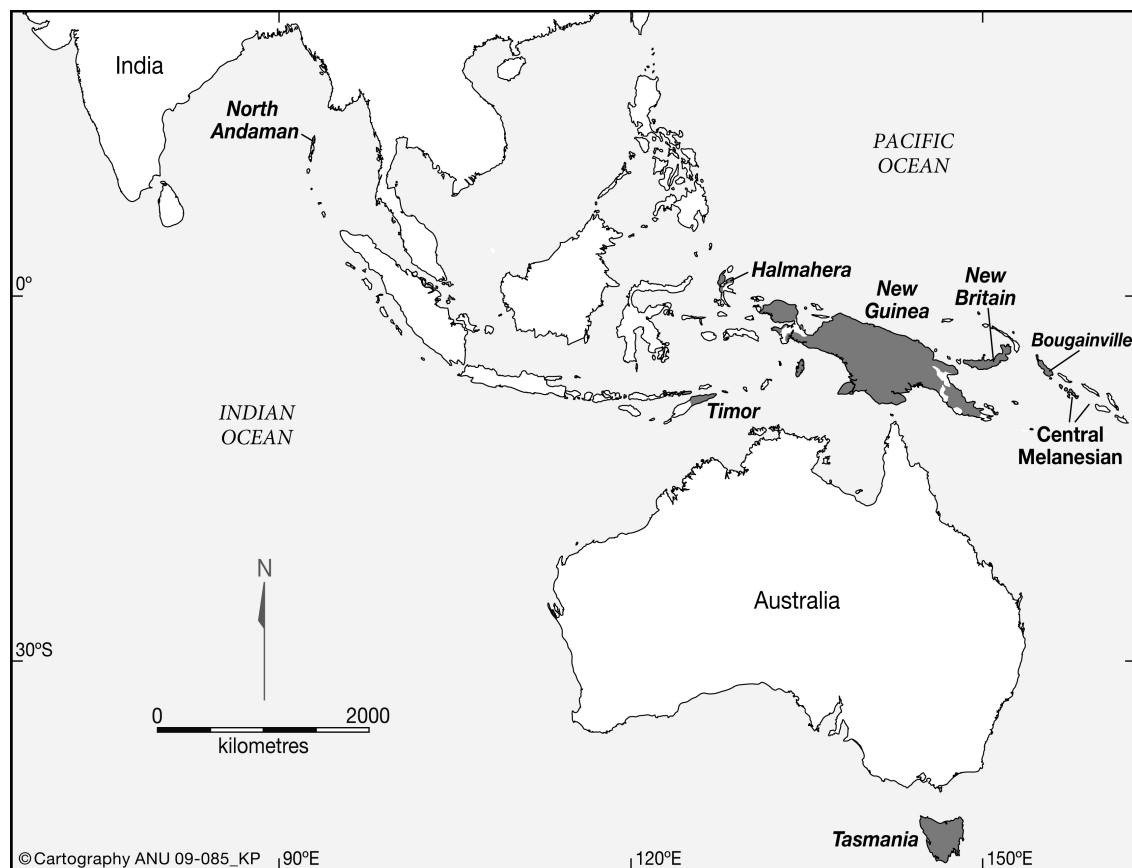
Twelve of these 14 groups are found in a more or less continuous area of the equatorial zone that extends from 123 degrees E to 166 degrees E, from Timor and Halmahera in the west, through the large (2400 km long) island of New Guinea, to New Britain, New Ireland and the Solomon Islands in the east. The dominant language family in this area, except for New Guinea, is Austronesian. The non-Austronesian languages, numbering between 700 and 800 (Wurm 1982; Wurm and Hattori 1981–83), are often collectively termed 'Papuan', traditionally without any implication that this label refers to a genetic grouping. The Andaman Islands lie much further west, south of Burma and the Bay of Bengal, at around 92 degrees E. Tasmania, situated to the south of the Australian mainland, is separated from all the other groups by a vast distance. Recently, Whitehouse et al. (2004) have argued that the Kusunda language of Nepal should be added to Indo-Pacific.

In this paper I will try to do four main things: (1) assess the merits of Greenberg's arguments for the Indo-Pacific hypothesis, (2) point out why specialists have up till now largely shirked this task, (3) evaluate weaker alternatives to the full-scale hypothesis, in which only some of the putative primary subgroups are included, and (4) reflect on the circumstances and chain of reasoning which persuaded Greenberg that he had a fairly good

¹ It gives me great pleasure to contribute to this volume honouring Malcolm Ross. It is my privilege and good fortune over the last 20 years to have had Malcolm as a colleague and friend and as a partner in a number of research projects in Austronesian and Papuan historical linguistics. I remain in awe of his accomplishments. Roger Blench, Beth Evans, Judith Blevins, Edgar Suter and Matthew Spriggs provided valuable comments on a draft of this paper.

case for a hypothesis when contemporary specialists in Papuan historical linguistics find these arguments unconvincing.

Beginning in the mid-1950s, Greenberg spent much time over a dozen years compiling putative Indo-Pacific etymologies, patiently tracking down data from obscure published and unpublished sources and entering materials in a notebook using his ‘multilateral’ or ‘mass comparison’ method, to be discussed in §5 below. In addition to the 84 sets of putative Indo-Pacific cognates he put together hundreds of other sets of resemblant words restricted to the putative subgroups of Indo-Pacific. Tentative findings were first reported in two unpublished papers, Greenberg (1958), where the 14 groups were defined and Greenberg (1960), where the full-scale Indo-Pacific hypothesis was first proposed. He continued to add to his materials until 1968, when he submitted the paper that appeared in 1971. According to Croft (2005:xviii), he examined some 350 lexical entries plus grammatical comparisons for about 800 non-Austronesian languages (plus some 50 neighbouring An languages for controls).²



Map 1: Location of the putative major subgroups of Indo-Pacific
(See §2 for details of the 14 groups)

² While Greenberg’s diligence in tracking down data was extraordinary, I doubt if he could have obtained 350 items for as many as 800 non-Austronesian languages in the 1950s and 1960s. There are probably fewer than 800 distinct Papuan languages in all, and for perhaps 200 of these there was almost no documentation during that period. However, Greenberg sometimes had wordlists for multiple dialects of single languages.

Greenberg (1971:854) writes that:

I believe that the evidence presented here is sufficient to establish the point that [in the Indo-Pacific region] the vast majority of non-Austronesian languages outside of Australia, on which judgment is still reserved, have a common origin.... For Tasmanian the relative paucity of data produces a somewhat weaker case than in other instances. Still what evidence we have does point in this direction.

He adds that 'My hope is that the present study will help to hasten the long overdue demise of the notion of Papuan as merely a scrapheap of assorted languages bound together by the negative characteristic of being non-Austronesian. May the comparative study of this major linguistic stock, which has been so strangely neglected, ... finally come into its own.' (1971:854)

The Indo-Pacific hypothesis has had a rather curious fate. Greenberg's arguments for Indo-Pacific have been summarily dismissed by various specialists as unconvincing, but none of the critics has provided a detailed assessment of the evidence. At the same time, the Indo-Pacific grouping is often mentioned in handbooks and encyclopaedias describing language families of the world, and is sometimes presented there as a more or less established stock (e.g. Ruhlen 1991). From time to time linguists doing comparative typological work (e.g. Viberg 1984) cite the hypothesis as if it were reasonably well supported, as do scholars in other disciplines, including population geneticists (e.g. Cavalli-Sforza 2000; Thangaraj et al. 2003) and historians (Manning 2006). This degree of acceptance is certainly not based on a rigorous assessment of the evidence — for no such assessment has been provided so far — but is surely due to the eminence of Greenberg himself, whose groundbreaking work on linguistic universals and on the classification of African languages made him one of the most influential linguists of the 20th century.

In recent decades research in archaeology and population genetics has greatly advanced our knowledge of the history of human settlement of Island SE Asia and the Pacific Islands. At present the most widely supported view among population geneticists and archaeologists is that the first successful colonisation of Asia, beyond the Levant, by *Homo sapiens* did not occur until between 70,000 and 50,000 years ago (Mellars 2006a, b). The Andaman Islanders, genetically, represent a clade of modern humans with no close relatives elsewhere (Thangaraj et al. 2003; Thangaraj et al. 2006). They appear to be a long isolated population deriving from the first modern human colonisation of South and Southeast Asia.³

It is now known that by at least 45,000 years ago modern humans were in Borneo, then still part of mainland SE Asia (Barker et al. 2005). By 45 to 40 millennia ago (and possibly several millennia before that) they had crossed Wallacea and reached Sahul, the Australia-New Guinea continent (Groube et al. 1986; O'Connell and Allen 2004; O'Connor 2007). Indeed as early as 40,000 BP people had made the sea crossings to New Britain and New Ireland (Leavesley and Chappell 2004; Pavlides and Gosden 1994; Specht 2005; Torrence et al. 2004) and by about 30,000 BP they were in Bougainville (Specht 2005; Spriggs 1997; Wickler and Spriggs 1988). The spread southwards across Australia, probably initially following the coasts, was quite rapid. By 35,000 BP the remote southwest corner of what is now the island of Tasmania was populated (Cosgrove, Allen and Marshall 1990; Mulvaney and Kamminga 1999). The genetic evidence indicates that, aside from some input from Austronesian speakers from SE Asia within the last three millennia or so, the

³ At present, archaeological dates for the Andaman Islands go back no more than about 2000 years but comparatively little archaeological work has been done there.

current Papuan-speaking peoples of New Guinea and Island Melanesia derive from these foundation populations (Friedlaender 2007).

Thus, if there was a common ancestor shared by the languages of the Andaman Islands, the Papuan languages of Melanesia, and the languages of Tasmania it is likely to have been more than 40,000 years in the past. If there was a common ancestral language shared by all the Papuan languages of New Guinea, the Bismarcks and the Solomons it was probably more than 30,000 years ago.⁴ Genetic and archaeological evidence indicate that there was little if any interaction between the founder populations of the Solomons and the rest of Melanesia in the period between initial settlement of the Solomons archipelago some 30 millennia ago (Friedlaender 2007; Friedlaender et al. 2008; Spriggs 1997) and the arrival of Austronesian speakers.

These chronologies do not rule out the possibility that the Indo-Pacific hypothesis is correct. It may be that the first early modern human expansion into SE Asia and Sahul was carried by very small populations speaking languages of a single family. However, the issue is not whether the Indo-Pacific languages share a remote common ancestor — indeed, all human languages may do so — but whether there is reasonable linguistic proof of common origin. The chronology for the first colonisation of Australia and Melanesia raises questions about what kinds of shared linguistic residues, if any, are likely to have survived after 40 millennia. From what we know of rates of replacement of particular kinds of lexical and grammatical roots, only a few dozen words have half-lives⁵ of more than 2000 years and the only elements that have half-lives of more than 20,000 years are some personal pronouns and a handful of lexical items, such as words for certain body-parts and kinship terms, and a few other concepts, probably fewer than 20 in all.⁶

It is unlikely that any cognate sets for items outside of the small hard core would have survived for 40 millennia, and even if they did, phonological changes would very likely have obscured their common origin. In this connection the fate of Greenberg's (1987) Amerind hypothesis (apparently independently developed in the 1950s by Morris Swadesh and Sydney Lamb) is instructive. Greenberg assigned to Amerind all 60 or so established stocks of Native American languages other than Eskimo-Aleut and Athabaskan.

⁴ These early dates were not known to Greenberg when he was formulating the Indo-Pacific hypothesis. Possibly he would not have been influenced by them because he believed in treating comparative linguistic evidence independently of non-linguistic evidence (Croft 2005:xii).

⁵ To say that the half-life of a word (more exactly, a particular lexical form-meaning pairing) is 2,000 years means that, in any language that has that it, the word has a 50 per cent chance of persisting (with the same meaning) for that period of time.

⁶ For a number of language families it has been shown that words for certain concepts are extremely persistent while words for other concepts are less stable. For discussion of Indo-European basic vocabulary see, e.g. Kruskal et al. (1971), Pagel (2000), Pagel and Meade (2006), Pagel, Atkinson and Meade (2007). Dyen et al. (1967) determined the cognation rates of words for 196 meanings in 89 Austronesian languages. Only 10 meanings show cognation rates of above 50 percent: *two, four, give, eye, we, louse, father, mother, to die, to eat*. Another 15 meanings show rates between 50 and 30 percent: *one, three, ashes, stone, nose, to hear, to drink, new, thou, ye, fruit, name, ear, liver, tree*. Another 45 show cognation rates of between 29 and 10 percent. Rates for the remaining 126 meanings fall below 10 percent. Austronesian is a family that lends itself well to determining absolute, as well as relative replacement rates in basic vocabulary. The dates at which Proto Austronesian and its major interstages were spoken are reasonably well-established because the Austronesian expansion left a well-marked archaeological trail (Bellwood 1997; Green 2003; Kirch 2000; Pawley 2002). This allows approximate retention rates to be calculated for particular lexical items reconstructed for Proto Austronesian (spoken about 5000 BP), Proto Malayo-Polynesian (about 4000 BP) and Proto Oceanic 3400–3100 BP).

Archaeology and genetics indicate that the Americas were probably not settled until about 15,000 years ago, and if these first settlers brought a single language one would expect the residue recoverable from comparison of the several hundred Amerind languages to be considerably larger than the residue recoverable for Indo-Pacific. Yet it has proved very difficult to persuade specialists that Amerind is a valid genetic stock (Campbell 1988, 1997; Campbell and Poser 2008; Mithun 1999; Nichols and Peterson 1996). It is not that the specialists are being obtuse but rather that their standards of proof are more demanding than Greenberg's were.

To estimate probabilities of chance resemblances ideally one needs data on the frequency of particular sounds in particular positions in all the relevant languages (Ringe 1992, 1996, 1999), data that are not available for Amerind. Instead, one must make do with approximations based on averaging data for certain languages. Following this procedure, Ringe (1996:152) 'finds no evidence whatsoever that the putative cognate sets in Greenberg's 'Amerind Etymological dictionary' represent anything other than chance'.

Although I will conclude that there is no good case for the full-scale Indo-Pacific hypothesis, Greenberg deserves credit for seeing, as early as the 1950s, that many of the small, disparate groups of non-Austronesian languages in the New Guinea area are probably related and, above all, for assembling a body of resemblant items that at least provide a basis for discussion. The material assembled by Greenberg includes some lexical and grammatical resemblances that indicate a common origin of most of the languages in six of the 14 primary groups that he posited. Subsequent work has shown that these languages belong to the language family now generally termed 'Trans New Guinea' (TNG). With some 400–450 member languages TNG is probably the third most numerous family in the world, after Niger-Congo and Austronesian.

However, Greenberg's failure to recognize that the TNG languages collectively amount to a single first-order witness rather than six severely weakens his arguments for the Indo-Pacific hypothesis as a whole. It turns out that by far the best evidence for Indo-Pacific consists of agreements among diverse branches of TNG. By contrast, the case for a genetic relationship between the North Andaman languages and the Tasmanian languages, on the one hand, and any of the Papuan groups of New Guinea, New Britain, Bougainville and the Solomons is extremely weak, the quantity and quality of the resemblances falling well within the range of chance.

Greenberg's material contains a few items that hint at a remote genetic relationship between the TNG family and certain other language families of New Guinea, and a similar connection may exist between certain non-TNG families of New Guinea and certain languages of New Britain. However, these do not amount to anything like a convincing case.⁷

2 Greenberg's subgrouping of Indo-Pacific languages

Greenberg's assumptions about the internal relationships of Indo-Pacific languages were central to his weighting of agreements in lexicon and grammar. He distinguished 14 major subgroups, which were assumed to have equal status as putative primary branches of Indo-Pacific. Most groups are defined by particular group-specific lexical resemblances (i.e. resemblances that go beyond the 84 putative Indo-Pacific etyma) and in some cases by

⁷ For more recent discussion of evidence concerning distant relationships among the diverse Papuan families see Dunn et al. (2002), Dunn et al. (2005), Reesink (2005).

particular grammatical features. A few of the groups were based on geography: they subsume different genetic stocks found in the same small region, which as a precautionary measure Greenberg treated as a single unit. Within most of the primary subgroups he distinguished further branches.

For some of the primary and secondary subgroups Greenberg lays out the evidence. In other cases he simply refers to published or unpublished evidence without citing details of resemblances. He offered words of caution:

This subgrouping is not exhaustive and is in some respects at least quite tentative ...

Such a degree of uncertainty is only reasonable at this stage' (1971:809).

Greenberg's subgroups, with the names and abbreviations he uses plus brief notes on the extent of the evidence he cites, are listed below. The order follows a directional pattern, moving roughly from west to east in the tropical zone, and then to Tasmania in the south.

The Andaman Islands

1. Andaman (AN). Greenberg observes that the Andaman languages fall into two groups that have not been shown to be related. He includes in Indo-Pacific only the larger, North Andaman group, which occupies almost all of Great Andaman Island and which consists of two closely related dialect clusters. He reserves judgment on whether the two South Andaman languages are related to the North Andaman group.

Indonesian archipelago

2. Timor and Alor (TA). Greenberg had data for only four of the non-Austronesian languages in the Timor-Alor region, at the eastern end of the Lesser Sundas chain, namely Abui, Bunak, Makasai and Oirata. Ninety-two comparisons are given in support of this grouping, of which 16 are also part of the list of 84 Indo-Pacific etymologies. He recognised two branches within the Timor-Alor group.

3. Halmahera (HA). The Papuan languages of north Halmahera 'form an obvious group ... so that no demonstration is necessary' (Greenberg 1971:815). They divide into a southern group, made up of Ternate and Tidore, and a northern group consisting of some 10 languages, including Loloda and Tobelo.

New Guinea mainland

4. West New Guinea (WNG). About 40 languages are named in this group group, all located at the western end of New Guinea, in parts of the Bird's Head and the Bomberai Peninsula. They are divided into four subgroups: (1) a large group of almost 20 languages including Etna Bay and Mairasi, (2) a group of about 10 languages including Madi, Tehit and Waken, (3) a group of four languages including Maibrat, and (4) Kapaur, Baha and Kovas. Twenty-seven etymologies were cited linking Etna Bay with Cowan's (1957) West New Guinea family. (More problematically, Cowan also included Halmahera and Timor languages in his West New Guinea family.)

5. Southwest New Guinea (SWNG) or Marind-Ok. Five subgroups are distinguished. Four of these, Tirio, Marind, Ok, and Awyu are located in south-central New Guinea, close to the Papua New Guinea border. The fifth, Kukukuku, located in Gulf and Morobe provinces, is a very tentative inclusion. About 60 supporting lexical comparisons are cited.

6. Southern New Guinea (SNG) or Kiwaic. In this group of some 50 languages Greenberg distinguishes seven branches, all spoken on or near the coast of southern New Guinea, in Western Province and the Torres Straits and on Frederik Hendrik Island. The subgroups are not given names but the largest are those that contain Kiwai and its immediate relatives, and Jei and its immediate relatives. Miriam, of the Eastern Torres Straits, belongs to SNG. The westernmost group is centred on Fredrerik Henrik Island. Seventy lexical comparisons support the putative SNG group.

7. North New Guinea (NNG). Under this rubric Greenberg combines several very divergent groups, scattered around north-central New Guinea on both sides of the West Papua-PNG border and in the Sepik provinces. He identifies these groups as Sentani, Tami, Arapesh, Murik, Monumbo (Manambu) and Ndu-Kwoma, along with some isolates. Fifty lexical comparisons plus pronominal agreements are cited in support of NNG.

8. Central New Guinea (CNG). This is the largest of the 14 groups assigned to Indo-Pacific. Greenberg, influenced by Wurm (1964), who had tentatively posited a large 'Central and Northeast New Guinea Phylum', recognised three primary branches: (i) Kapauku-Baliem (also known as Ekagi-Dani) in the western highlands of West Papua, (ii) a Central Highlands group (similar to Wurm's East New Guinea Highlands Stock) and (iii) Huon (in the region of the Huon Peninsula, just north and east of the Central Highlands group). The Central Highlands group in turn divides into several groups, including those known nowadays as Engan (including Huli, Mendi, Kewa and Ipili), Chimbu-Wahgi, and Kainantu-Gorokan. No supporting cognate sets are cited other than those in the main Indo-Pacific list. For the smaller groups he refers to the published work of others.

9. Northeast New Guinea (NENG) or Madang. To this Greenberg assigns 30 or so languages of Madang Province. He remarks that the nucleus of such a group was recognised by Ray (1919) and that '[t]he unity of this group is quite obvious' (1971:834). No etymologies are cited.

10. East New Guinea (ENG). This putative group has more than 80 members located in the southeastern region of New Guinea. Thirty-seven lexical comparisons and four pronominal agreements are cited in support of ENG. Greenberg recognises ten subgroups, including the larger groups he calls Mailu, Binandere, and Koita, along with Dimpa, Kovio, and Elema.

The Bismarck Archipelago

11. New Britain (NB). Greenberg had data for five languages, all of Central and East New Britain. He recognised that there are two or more very divergent groups on this large (450 km long) and mountainous island, prone to cataclysmic volcanic eruptions, but treats them as a unit for geographic reasons. Four languages, comprising the Baining-Sulka group, are clearly related. He was uncertain whether the fifth, Uasi, had any special relationship to other New Britain languages and he had no data on a language, Idne, said to be spoken in the far west of the island. No etymologies are cited in support of NB.

The Solomon Islands

12. Bougainville (BO). There are two highly divergent groups, one in the north and one in the south. Each contains four languages (some with diverse dialects). Greenberg considers that they have enough in common to justify a Bougainville subgroup (2005:203) but he cites no cognate sets in support of this claim.

13. Central Melanesia (CM). There are four non-AN languages of the central Solomons ‘which seem to constitute another subgroup’ (Greenberg 1971:816) that he calls Central Melanesian. He also tentatively assigned to this group certain languages of Reefs/Santa Cruz Islands, situated almost 1000 km to the east, mainly on grounds of resemblances in the pronouns. Fifty-two lexical and nine pronominal etymologies are cited for CM.

Tasmania

14. Tasmania (TS). The only data consist of brief and highly problematic notes made by 19th century colonials. Following Schmidt (1952) Greenberg recognises five languages (or dialects): a northern language and four others that appear to be more closely related to each other. No supporting lexical data are cited, other than those in the Indo-Pacific list.

Following Cowan (1957, 1960), Greenberg (1971:839) speculates that Halmahera, Timor–Alor and West New Guinea may constitute a ‘supergroup’, on the basis of some agreements in grammatical features.

Unclassified languages of New Guinea (UNG). Greenberg also referred to, but left unclassified, a number of very small groups and isolates in New Guinea and to one isolate spoken on New Ireland.

3 The reception of the Indo-Pacific hypothesis. Why have specialists largely ignored it?

Why has the Indo-Pacific hypothesis received little attention from specialists in the relevant language groups? In the four decades since Greenberg’s main publication on this subject there have been a handful of brief assessments by specialists, consisting of just a few sentences, and all have rejected the evidence as unconvincing (among these are Laycock 1975a; Pawley 1998, 2005a; Ross 2005). It is noteworthy that in Foley (1986), a book devoted to the Papuan languages, there is no reference to the Indo-Pacific hypothesis and that in another book on the Papuan languages, Wurm (1982:6, 30) simply acknowledges Greenberg’s proposal in three sentences.⁸ There have been a couple of papers that briefly examine Greenberg’s arguments for particular subgroups of Indo-Pacific (Franklin 1973; Voorhoeve 1975). This reception stands in contrast to the lively and extensive debates generated by Greenberg’s African and Amerind classifications.

As far as I know, only one slightly more extended review of on the Indo-Pacific hypothesis has appeared: a five page commentary by Wurm (1975a:925–929). However, this commentary contains much hedging and little discussion of nitty-gritty details. The main points could have been made in half a page. They are that:

- (i) Greenberg made several claims about relationship between diverse Papuan groups that now, in the light of better data than he had, appear not to be demonstrably related. This in turn casts serious doubt on the value of his evidence for the claim that Tasmanian and Andaman are also related to the Papuan languages.

⁸ I suspect that neither Wurm nor Foley wished to offend a respected colleague and chose not to air their disagreements. Foley in particular has close links to Stanford.

- (ii) The case for Tasmanian is particularly weak. Among the grammatical items, there is a single pronominal resemblance, in the 2SG forms. Eighteen of the 84 lexical etymologies include Tasmanian items but these are not convincing.
- (iii) The number of resemblances exhibited by the Andaman group seems, at first blush, to be significantly higher. There are three pronoun items that show a resemblance to pronouns found in certain other groups, plus the past tense marker *k*. But Wurm observes that the pronominal agreements are really much weaker than the foregoing statement implies, because they are divided among disparate groups: the 1SG agreement is with 'West Papuan', the 2SG agreement is with 'East Papuan', and the 1PL agreement is with yet another set of languages. Thirty of the 84 Indo-Pacific etymologies are represented in Andaman, and the resemblances are chiefly with W. Papuan and Timor-Alor languages. Wurm suggests that there may be an ancient substratum in the Papuan area that is linked to the Andaman languages.

There are, I believe, several reasons why scholars have been reluctant to attempt a detailed assessment of Greenberg's Indo-Pacific proposal. First, a thorough review would be very time-consuming. The Indo-Pacific hypothesis is in fact a cluster of many hypotheses about genetic groupings, each of which has more or less independent status and would have to be assessed one by one. Apart from the full-scale Indo-Pacific grouping there are the 14 proposed subgroups, many of which are highly problematic.

A second reason has already been alluded to above: Greenberg did not separate the wheat from the chaff. He had the makings of a good case for linking several Papuan stocks in New Guinea with each other and with certain languages of the Timor-Alor area, but did not separate this from the much flimsier case for including the languages of Tasmania, the Andamans, Halmahera and Island Melanesia. Unsurprisingly, some readers faced with these more far-reaching and weakly supported claims, were inclined to ignore the rest.

Third, the pool of specialists who are more or less competent to review the evidence has always been very small. The total number of linguists actively working on the historical study of any or all of the Papuan families has probably never exceeded ten or twelve at any one time (the peak was between about 1965 and 1975) and since the early 1980s has been considerably fewer. Among these scholars, only one or two have had Papuan historical linguistics as their primary research field. As for the Andaman and Tasmanian languages, the situation is worse.

Fourth, the timing of Greenberg (1971) was unlucky. His thunder was largely stolen by discoveries in Papuan studies that were being reported in the 1960s and 1970s. The idea that some of the diverse, small Papuan groups might be related was in the air during the 1950s, as can be seen in the writings of Capell (1948–49), Cowan (1957), Loukotka (1957) and Wurm (1954). However, the kinds of arguments put forward in these works were chiefly typological, and we can now see that in many cases the early tentative proposals did not stand up.

Beginning in the late 1950s, Stephen Wurm, at The Australian National University, initiated a long term program of field surveys and comparative research on the Papuan languages of New Guinea and Island Melanesia. He was soon joined by several collaborators — both departmental colleagues and PhD students. In the early 1960s Wurm published a series of papers giving typological and lexicostatistical evidence for a family of around 50 languages located in the central highlands of Papua New Guinea, which he called the East New Guinea Highlands Stock (later called a Phylum) (Wurm 1960, 1964,

1965). In the mid-1960s several proposals concerning long-range connections among diverse groups in New Guinea were put forward, such as Wurm's (1965, 1971) Central New Guinea Macro-Phylum.⁹ These were based mainly on typological agreements, the lexicostatistical agreements between widely separated groups being too low (2–5%) to be significant.

Then at the end of the decade McElhanon and Voorhoeve (1970) cited about 90 sets of resemblant lexical items shared by several widely separated groups: namely a group located in central and south-central New Guinea (Voorhoeve 1968), another in the Huon Peninsula area of central north New Guinea (McElhanon 1967, 1970) and a third, the small Binandere family, situated in the southeast of the island. They coined the name 'Trans New Guinea phylum' for this widespread group. McElhanon and Voorhoeve used a method very similar to Greenberg's multilateral method (§5.1) to assemble possible cognates, without attempting to work out regular sound correspondences. Among their sets of resemblant forms were several pronouns and other items of core basic vocabulary.

This first, restricted version of the TNG hypothesis was soon dramatically extended — indeed McElhanon and Voorhoeve (1970) had suggested that their TNG Phylum would turn out to be related to the groups in Wurm's Central New Guinea Macro-Phylum. Within a few years, the central highlands family and various other groups, including the Timor area languages, had been added to TNG, so that almost 500 languages, or about 70 percent of all non-Austronesian languages of the region, were assigned to this family in its most extended form (Wurm ed. 1975; Wurm 1982; Wurm et al. 1975).

As it happens, proponents of the extended TNG hypothesis in the 1970s did not make good use of the evidence they had. The case they made for TNG was poorly made and far from convincing because it relied too much on typological resemblances, and provided no systematic phonological and lexical reconstruction. All informed reviewers were highly sceptical (Foley 1986; Haiman 1979; Heeschen 1978; Lang 1976). However, more recent work has yielded more solid evidence for TNG (with a membership not quite as extensive as that proposed in Wurm (ed. 1975)). Accounts of the history of the TNG hypothesis are given in Pawley (1998, 2005a) and Ross (2005).

In the 1970s several other major genetic groupings besides TNG were posited by the ANU group. These included the Sepik-Ramu Phylum, to which were assigned almost 100 languages of north central New Guinea (Laycock and Z'graggen 1975), the Torricelli Phylum, consisting of some 47 languages of the Torricelli Ranges and nearby regions of the Sepik and Ramu Provinces of Papua New Guinea (Laycock 1975b) and the East Papuan Phylum, said to subsume all 20 or so non-Austronesian languages of Island Melanesia (Wurm 1975b). Recent opinion is that neither the Sepik-Ramu nor the East Papuan groupings stand up, the evidence for Sepik-Ramu being flimsy (Foley 2005) and that for East Papuan even more so (Ross 2001). However, these speculative groupings were included in the influential *Atlas of languages of the Pacific* (Wurm and Hattori 1981–83) and outsiders have often assumed that they are well-supported genetic groups.

⁹ The extended TNG hypothesis had in fact been roughly foreshadowed in a 1965 report by the Voegelins, where they proposed to unite the East New Guinea Highlands Stock with a Huon Peninsula group, the Binandere group, the Ok group of central New Guinea, and the Dani group of the SW New Guinea Highlands. They also threw in the Ndu family of the Sepik, which is not TNG. However, the Voegelins relied on the data and proposals of Greenberg, Wurm and other sources and did not analyse the evidence further.

Finally (and this is arguably the most important single reason for the lack of a detailed assessment) there are major methodological difficulties in evaluating the evidence. Greenberg's method of multilateral comparison yielded a rather small body of impressionistic resemblances between form-meaning units. In such cases, unlike claims about cognation that rest on regular sound correspondences, one cannot appeal to rigorous and reliable criteria to assess a claim of common origin. The claims crave statistical testing for significance but such testing would be time-consuming and few linguists are well equipped to do it.

Such methodological concerns recur in most of the brief assessments of Greenberg (1971), for example, that of Laycock (1975a:57):¹⁰

To date it can safely be said that there is no real evidence to link the [non-Austronesian] languages of New Guinea with any other linguistic groups ... In particular Greenberg's Indo-Pacific hypothesis ... is not only far from proven, but also based on inadequate and insufficiently analysed data (for example, comparisons are too frequently made of items within larger groups of languages — such as the Trans New Guinea Phylum that are already known to be related, so that there is little support for the wider relationships postulated.)

4 Contemporary views of the genetic classification of Papuan languages

Contemporary views of the classification of the non-Austronesian languages of Melanesia and Southeast Asia, and the languages of Tasmania, differ in various ways from Greenberg's. A detailed account of contemporary views would require a separate paper. A thorough review of the Tasmanian data was undertaken by Crowley and Dixon (1981). Research on the Andaman languages is reviewed by Blevins (2007, to appear). Wurm (1975a, 1982) surveyed work on the 'Papuan' languages up to the mid 1970s. The most comprehensive recent classification of the 'Papuan' languages, based mainly on pronominal paradigms, is in an unpublished paper of some 200 pages by Ross (2000), whose main findings are summarised in Ross (2005) and in Pawley (2005b, 2007). Among other studies that treat some of the groups accepted or proposed by Greenberg are the following. Z'graggen, in a number of works (e.g. Z'graggen 1975), confirmed and extended the Northeast New Guinea (Madang) group. Ross (2001) examined Wurm's (1982) hypothesis that the various non-Austronesian families of Island Melanesia (Melanesia excluding New Guinea) belong to a diverse East Papuan phylum. Foley (2005) argued against the Sepik-Ramu hypothesis (the core of Greenberg's North New Guinea group). Pawley (1998, 2005a, b) and Ross (1995, 2005) argue in support of the Trans New Guinea hypothesis. Voorhoeve (2005) discusses inheritance and diffusion among certain of the groups making up Greenberg's Southwest New Guinea group and Reesink (2005) does the same for West New Guinea. Dunn et al. (2002, 2005) examine the distribution of a large selection of typological characteristics across the various families that Wurm (1982) had assigned to 'East Papuan', in an attempt to find traces of ancient common origin or diffusion.

¹⁰ However, the words 'already known to be related' in this quote are unfair to Greenberg. The 'larger groups' that Laycock refers to, such as the Trans New Guinea Phylum and Sepik-Ramu Phylum, were not proposed in print until the early 1970s and even then were not well supported (see below). Greenberg (1971) was submitted in 1968, some three years before it was published as part of a large multi-authored volume. I am confident of this, first, because none of the chapters in that volume contain references dated later than 1968 and because some other contributors to this volume told me they had a deadline of 1967 or 1968.

The most important differences with Greenberg's views concerning genetic relationships to emerge are listed below:

- (i) It is now clear that (leaving aside certain problematic single languages) almost all the languages in Greenberg's groups 5 (Southwest New Guinea or Marind-Ok), 8 (Central New Guinea), 9 (Northeast New Guinea or Madang), 10 (East New Guinea), and part of group 6 (Southern New Guinea or Kiwaic) belong to a single large family, Trans New Guinea (Pawley 1998, 2001, 2005a, b; Ross 1995, 2000, 2005). Of particular importance is the fact that the non-Austronesian languages of group 2, Timor and Alor (along with those of Pantar), geographically isolated from the rest, have fairly strong claims to be assigned to TNG. Indeed, Ross (2000), on somewhat slender pronominal evidence, specifically assigns them to a subgroup that has other members on the New Guinea mainland, around the Bomberai Peninsula.
- (ii) Although all the languages Greenberg assigned to group 5, Southwest New Guinea (or Marind-Ok) belong to TNG, this set of languages is not now regarded as forming a subgroup (Pawley 2005a; Voorhoeve 2005).
- (iii) Group 6, Southern New Guinea, is not regarded as a genetic group but is divided into several families, one of which is TNG. Ross (2000) tentatively includes Kiwai and its immediate relatives in TNG but not the rest of group 6.
- (iv) Group 7, Northern New Guinea, is not regarded as a genetic group but is divided into several families and a few isolates (Foley 2005; Ross 2000). A few of the languages Greenberg assigned to NNG are TNG.
- (v) Although all the languages Greenberg assigned to group 8, Central New Guinea, are now assigned to TNG, they are not viewed as otherwise forming a subgroup. Indeed, the large Central Highlands branch of CNG posited by Greenberg (following Wurm) is not now regarded as forming a subgroup of TNG. On the contrary, the Central Highlands languages fall into several groups that on present evidence appear to be first-order branches of TNG.
- (vi) Although all the languages Greenberg assigned to group 10, East New Guinea, belong to TNG, they are not now viewed as forming a subgroup.
- (vi) Group 11, comprising the New Britain languages, divides into at least two families (Ross 2000, 2001), a possibility that Greenberg acknowledged.
- (vii) Group 12, Bougainville. Ross (2000, 2001) finds no case, on the pronominal evidence, for uniting the two highly divergent groups, one in the north and one in the south.
- (viii) Group 13, Central Melanesian. Ross (2000, 2001) finds very weak evidence for relating the four non-AN languages of the central Solomons. Ross and Næss (2007) have shown that the Reef Islands language is not 'Papuan'. It belongs to the Oceanic subgroup of Austronesian and, by association this holds for the Santa Cruz languages, which are its immediate relatives.
- (ix) Group 14. Crowley and Dixon (1981) conclude that there were at least six distinct languages represented in the meagre data recorded from Tasmania, but probably between eight and twelve. The materials consists of 200 to 300 words for some South-east lects and much smaller amounts for other lects. The only clear grammatical data available are forms for 'I' and 'you' in a few languages.

Although some of the languages are clearly related, the data do not permit the conclusion that all the Tasmanian languages are related. And,

[a]lthough Tasmanian languages seem typologically similar to languages of the Australian family [in their phonologies], there are insufficient cognates [read 'resemblant forms'] to justify an even tentative hypothesis of genetic relationship (Crowley and Dixon 1981:395).

5 On the lexical evidence for Indo-Pacific

5.1 Greenberg's etymologies

In various places in his writings Greenberg makes the point that the first step in the comparative method is working out which languages to compare, i.e. which languages are likely to be genetically related. He regarded his method of 'multilateral comparison' as an efficient way of carrying out this first step. He described the method as one that 'looks at everything at once' (2005:94). Word lists are arranged so that one's eye scans a few words across many languages, rather than many words across a few languages. That is, there is simultaneous comparison of languages and lexical items from the full range of languages and language families under consideration. Greenberg makes the following observation about the value of the method as a discovery procedure.

Most important of all, perhaps, is that where more than one family is represented, ... the contrast between the relatively numerous and qualitatively superior resemblances between related languages, compared to the sporadic and qualitatively poorer resemblances among unrelated languages, becomes readily apparent. In this way the presence of unrelated languages provides a control for distinguishing mere chance from genetically significant resemblances. (Greenberg 2005:42)

This observation is surely true but there is a certain irony in it, when we consider the quality of the evidence for Indo-Pacific.

As an example of the power of the multilateral method Greenberg lists words for diverse European languages, organised so that all the Germanic languages are contiguous, likewise the Celtic languages, the Romance languages, and so on, and writes

In Table 7 I have listed a few basic words for twenty-five languages of Europe. The number of ways of classifying twenty-five languages, even without specifying subgroupings, is 4639×10^{19} , that is, over a quintillion. Yet the correct classification and even subgroupings and intermediate groupings (e.g. Balto-Slavic) are apparent from just a cursory glance at two or three words (2005:94).

(One can accept Greenberg's main point here but it should be noted that he has organised the table to make this easy. A random listing would take more than a cursory glance to sort out.)

That is all very well for Indo-European and its major branches but it is clear multilateral comparison does not work so well when the groups are, at best, only very distantly related — otherwise, of course, there would not be such a level of disagreement among scholars as we find. Multilateral comparison relies on there being enough resemblant items shared by a pair (or larger set) of languages to decisively indicate common origin without the time-consuming work of establishing regular sound correspondences. For the putative high-order subgroups of Indo-Pacific we cannot compile tables comparable to those available for Indo-European because the number of resemblant forms in basic vocabulary is much, much smaller. And this is the critical difference. The problems are to know (a) what counts

as resemblant items, (b) how many such items are enough and (c) how to distinguish cognates from chance similarities and borrowings.

The pitfalls of trying assessing resemblances without knowing the phonological history of the languages are illustrated by #38 'head'. Greenberg compares CM languages that have forms of the type of Savo *mbatu* with Bunak (Timor) *ubul* and Yela Dne (Rossel Island, SE Papua New Guinea) *mbara*. But apart from the phonological differences, the CM forms are Austronesian loans: reflexes of Proto Oceanic *b^watu 'head' are widely reflected (as *mbatu*, etc.) in the Oceanic languages of the Solomon Islands.

Greenberg is wont to quote statistics indicating that the chances of certain sets of resemblances occurring by chance are infinitesimally remote. One must take these estimates with a large grain of salt, because all too often there are counterexamples. What are the chances that English and Maori, two unrelated languages, would show marked similarities in the numerals 2, 3, 4? They do. Compare English *two* (Scots *twa*), *three*, *four*, with Maori *rua*, *toru*, *whaa* (where *wh* is a bilabial fricative). All the Germanic languages show comparable likenesses to almost all the Polynesian languages.

As a sample of the difficulties posed by the proposed Indo-Pacific etymologies consider comparisons #56 to #59. In #56, for the meaning 'old', resemblant forms are cited from witnesses in four far-flung groups: Andaman (four languages) *tam* and *taum*, Halmahera (one language) *timono*, Central New Guinea (two languages) *tamana*, *tamon*, and the Solomon Islands (one language) *tam*. In #57, headed 'to plait', resemblant forms are cited from just two groups: Andaman (Biada *tepi*) and Halmahera (Tobelo *tapi*). In #58, for 'to push', forms are cited from two groups: Andaman (Bogijeb *tera*) and Halmahera (Tobelo *tila*). In #59, headed 'rain', forms are cited from four groups: Tasmania (four languages have *moka* 'water'), WNG (two languages have *moka* 'wet', NNG (seven languages have a range of forms such as *mayk*, *mac*), and SWNG (two languages have *mauka* 'water).

In the absence of any knowledge of the historical phonology of any of the languages cited, what can be said about these resemblant items? We can note the formal similarities and ask what is known about the stability of terms meaning 'old', 'to plait', 'to push', etc. We can speculate on how likely it is for a few languages out of 750 to retain such resemblant forms after 40,000 years of separate development, and try to calculate how likely it is that such resemblances could have developed independently ('by chance') in different groups. But without a rigorous statistical analysis using fair and reasonable criteria there is simply no way of separating the wheat from the chaff other than one's personal judgment.

I consider that, among Greenberg's 84 Indo-Pacific etymologies, about 23 contain a core of convincing resemblant items. A list of the most promising etymologies is given below. The sets are numbered as in Greenberg's list but I have greatly abbreviated the material. Instead of citing long lists of forms from those of Greenberg's subgroups that we now assign to TNG, I cite a reconstruction attributable to an early stage of TNG (here labelled simply 'pTNG'). In reconstructed forms C = consonant, V = indeterminate vowel. Most of the reconstructions are drawn from Pawley (2005a, n.d). Particular resemblant forms are cited from Andaman and Tasmania languages but for other non-TNG groups I merely note, in most cases, that a particular subgroup is represented in the set of resemblances. Putative resemblances cited by Greenberg that seem very far-fetched are discarded from the comparisons listed below.

Table 1: The most promising of Greenberg's Indo-Pacific etymologies

above (1)	PTNG *op(V)
arm (4)	PTNG *mbena, AN <i>ben</i> 'shoulder-blade'
bark (7)	PTNG *ka(nd,t)apu, AN <i>kait</i> , <i>kaic</i> , TS <i>kite</i> , NNG
bone (12)	pTNG *kondaC, TS <i>teni</i>
come (18)	PTNG *ma(n)-
die (21)	PTNG *kumV-
ear (23)	PTNG *damV, NB, NNG
earth (24)	PTNG *ma(l,n)a
eat (25)	pTNG *na-, BO, NNG
egg (26)	PTNG *mangV 'round, compact object', AN <i>molo</i> , <i>mula</i> , <i>mule</i>
female (28)	PTNG *pan(V), BO
fire (30)	PTNG *inda, AN <i>at</i> , TS <i>to</i> , <i>toi</i>
hair (33)	PTNG *iti, AN <i>de</i> , HA ??
husband/male (42)	PTNG *ambi
lip/mouth (45)	PTNG *ambe, AN <i>pe</i> , <i>pa</i>
louse (47)	PTNG *niman, NNG, NB, UG
moon (51)	PTNG *kal(a,i)m
nose (58)	PTNG *mundu
older sibling (63)	PTNG *nan(a,i)
stay (65)	PTNG *mVna-
star (71)	PTNG *bay, TS <i>poe</i> , ENG, CNG, UNG
stone (73)	PTNG *kambu(CV), BO, NNG
tongue (76)	PTNG *me[l,n]e, TS <i>mena</i> , BO, UNG

All or almost all of the remaining Indo-Pacific etymologies, close to three-quarters, can be discarded as 'chaff'. A good many of the putative cognate sets represent meaning-form pairings that typically have quite short half lives, e.g. 'arrow', 'beautiful', 'bush, forest', 'buttocks', 'to dance', 'fog', 'mud', 'to plait', 'thing' 'to push', 'to walk', 'white', 'yellow'. The fact that Greenberg is able to find roughly similar forms for these concepts in diverse Indo-Pacific groups, that have independent histories for the past 30 millennia, must weaken our confidence in the reliability of the method. A good many resemblances are only included by allowing the semantic net to be cast very wide. For example, the set of 'earth' includes forms glossed 'bottom', 'underneath', 'mud', 'land'; under 'walk', are included forms meaning 'leg, foot'; under 'ear' are included verbs 'to hear'; and so on.

My view is that none of the lexical resemblances between North Andaman and Tasmanian languages and between members of either of these groups and other groups assigned to Indo-Pacific are due to common origin. There are three reasons for this conclusion:

1. The Andaman and Tasmanian populations have been isolated from each other and from the Papuan speaking peoples of Melanesia for at least 40,000 years (see discussion in §1). Everything known about rates of lexical replacement in large language families indicates that the shared lexical residues left after 40 millennia are likely to be very, very meagre and entirely confined to a small core of basic vocabulary, probably fewer than 20 words. Furthermore, phonological changes would very likely have obscured the common origin of almost all the surviving cognates.

2. There is no compelling collection of resemblances in the hard core basic vocabulary. Only one or two noteworthy agreements are found in that domain: Tasmanian *mena* ‘tongue’, TNG *me[l,n]e, and perhaps AN *pe* ‘lip, TNG *ambe ‘mouth’. These isolated likenesses are not enough to make a case.
3. Given the very large number of languages compared the overall number of lexical resemblances is small and not above chance levels. Among these likenesses are some that are too good to be true — very similar forms for meanings that are not core basic vocabulary. There are superficially impressive resemblances between many language families that are not generally regarded as related. For instance, enthusiastic amateurs (and occasionally professionals) have come up with hundreds of look-alikes shared by Semitic and Austronesian, by Japanese and Austronesian, by Quechua and Austronesian, and even by Niger-Congo and Austronesian.

What about Greenberg’s lexical evidence for relating what we now know to be the TNG family to other putative Indo-Pacific subgroups from Bougainville and the Central Solomons? Essentially the same objections apply to this evidence as to the case for relating Andaman and Tasmanian languages to the languages traditionally known as Papuan. As noted earlier, the archaeological record suggests that, following initial settlement of what was then the island of Greater Bougainville some 30,000 years ago there was little or no contact between New Guinea populations and populations in Bougainville and the Solomons until the advent of Austronesian speakers around three millennia ago and the genetic record is consistent with this conclusion.

The lexical evidence for connecting TNG with certain other languages of the New Guinea mainland is, I think, slightly stronger. For example, forms resembling the very stable TNG etyma *niman ‘louse’ and *na- ‘to eat’ occur in a number of non-TNG languages of New Guinea. But there is no space here for a detailed assessment of this evidence.

6 On the grammatical evidence for Indo-Pacific

Greenberg (1971:842ff.) cites agreements in 11 grammatical features (and alludes to others). He regarded these, especially certain pronominal agreements, as the strongest part of his evidence for including the various non-New Guinea groups in Indo-Pacific. The trouble is, again, that the strongest agreements are between members of TNG. As recent work has confirmed, one can reconstruct for pTNG a complete paradigm of independent personal pronouns and part of a set of verbal suffixes marking subject person-and-number and some other fragments of morphology. The problem is to make a case for reconstructing grammatical features to a stage earlier than pTNG.

The following table of TNG independent pronouns is based on Ross (2005:29), as slightly modified in Pawley (2005a:89):

Table 2: Proto TNG independent pronouns^a

	1	2	3
singular	na	ŋga ^b	[y]a, ua
plural (i-grade)	ni	ŋgi ^b	?
plural (u-grade)	nu		
dual (i-grade)	ni(l,t)i	ŋgi(l,t)i ^b	i(l,t)i
dual (u-grade)	nu(l,t)i		
non-singular	nja		

^a Ross also reconstructs an inclusive suffix *-m- 'plural' and *-p- 'dual' (2005:29).

^b A case can be made for reconstructing the initial consonant as *k rather than *ŋg.

The following is a critical summary of Greenberg's account of the grammatical evidence for Indo-Pacific.

1. First person singular pronouns. He notes that two sets of forms are widespread.

- (i) *n*-forms 'absolute (independent)'. The reconstruction of pTNG *na '1SG independent' is generally accepted. This accounts for the occurrence of *n*-forms in TA, CNG, SWNG, SNG, NENG, and ENG. (Within TNG, *na reflexes are absent from the Madang and SE Papuan groups.)

Outside of TNG *n*-forms are found in:

West New Guinea: The Konda-Jahadu and Kapaur groups have *n*- (Kampong Baru *neri* '1SG', *eri* '2SG', Tarof *ne(iga)* '1SG' *va(iga)* '2SG').

North New Guinea: Ndu has *n*-forms (Maprik *unə*, Kwoma, Mayo *an*).

Bougainville: Telei *na*, Nasioi and Koromiva *n*- 'my'.

Central Melanesia: Savo *-ni* '1SG obj.', *n*- + '1 object marker'. (The 'cognates' in Santa Cruz languages, which are now classified as Austronesian, are invalid.)

- (ii) *t*-forms for subject and object. Within TNG, these are found in TA (Makasai *ani* '1SG absolute', *asi* possessive), Kainantu: Benabena *nani* absolute, *-te* possessive (and other Tairora group languages). In NENG *t*-forms are widespread for both absolute and possessive uses.

Outside of TNG, *t*-forms are found in WNG. About half of WNG languages have *t*-forms for subject and object, and the other half have *n*-forms.

2. Second person singular pronouns. Greenberg finds that 'over a large part of New Guinea' (1971:844) there is an opposition between first person *n* (usually *na*) and second person *k*- (usually *ka*). (Here he has recognised the TNG pattern. pTNG *ka or *ŋga '2SG' is well attested.) Where the *na/ka* pattern does not predominate the most common second person pronoun is *ngi* or *ni*. 'I suspect that *ngi* is original and has frequently become *ni* either by direct phonetic change or under the influence of first person singular *n*' (1971:844).

Beyond TNG, 2SG forms with initial *n* or *ng* occur in:

Andaman: Biada *ngol*, Onge *ngii*.

Halmahera: Galela *no* '2SG subject', *ni* '2SG object', *ngona* '2SG independent'.

West New Guinea: Amberbaken, Madik, Karon *nan*, etc.

North New Guinea: Tanggum *nu*, Murusapa *na*, Anaberg *nə*.

New Britain: Baining *ngi*, Taulil *nggi*, *ngginggi*, Uasi *nini*.

Central Melanesia: Savo *no*, Bilua *ngo*, Baniata *no*.

Tasmania: All dialects *ni(na)*.

3. First person plural pronouns. Greenberg found *ni* is widespread in groups that we now assign to TNG. pTNG **ni* and **nu* '1PL' are well attested.

Similar forms are found in certain other languages of New Guinea:

West New Guinea: *ni(ti)* '1pl.excl.' in Solowat, Higo and congeners.

North New Guinea: Anaberg *ni*, Tanggum *nai*, Sko, Sagke *ne*, Ndu *nanə*, etc.

Unclassified New Guinea: Rossel (Yela Dnyi) *nu*- 'our'.

Beyond New Guinea, *n*- initial forms are found in North Bougainville (Telei, Nasioi *nii*, *ni* 'our') and Halmahera (*na* marks 1pl object inclusive in most HA languages).

At first blush these resemblances in first and second person pronouns between TNG and members of other groups listed in (1–3) above look impressive. However, there are a number of grounds for caution. Ross (2005:50) is critical of Greenberg's application of the multilateral method to pronouns in his Amerind work, treating pronouns as individual forms rather than as part of paradigmatic sets. In his Indo-Pacific study Greenberg cites paradigms where possible but does not hesitate to include resemblant pronominal forms that come from different paradigmatic sets.

There is a general problem in evaluating formal resemblances among pronouns. Rhodes (1997) argues that functional pressures restrict the range of phonological features used to mark pronominal contrasts, thereby increasing the likelihood of chance similarities among pronoun forms. One such factor is that pronouns or pronominal affixes are typically backgrounded in discourse. This means, among other things, that they tend to be short (singular markers almost always a single syllable) and unstressed. Three problems must be solved for backgrounded items to be communicatively effective:

- (a) identification: one must be able to tell when one is hearing a morpheme of the relevant type, e.g. a pronoun, not a noun.
- (b) differentiation. One must be able to distinguish among members of this class.
- (c) ease of pronunciation. One must be able to pronounce the items with relative lack of attention.

These factors stand in partial conflict and produce a range of optimal pronominal systems.¹¹ The ease of pronunciation consideration strongly favours use of unmarked segments, i.e. the more common or most common segments in pairs or larger sets of phonemes. Rhodes cites work by Gordon (1995) who using a sample of 62 languages of diverse families found that consonants and vowels occur in pronominal systems with the following frequencies (percentages rounded out).

¹¹ Rhodes' account of the differentiation and identification problems refers to quite complex factors that allow a variety of optimal systems and I will say almost nothing about these here. The differentiation problem favours systems that maximize acoustic distinctness but not in a way that reflects any sound symbolic link between one of the persons and one of the classes of sounds.

Table 3: Frequencies of consonants and vowels in pronominal systems across 62 languages

consonant	% of languages	vowel	% of languages
n	93	a	98
m	75	i	90
k	71	u	69
t	68	o	56
y	53	e	52
w	43		
h	40		
ŋ	39		
s	37		
r	37		
...			
ñ	19		

That is to say, this factor favours the use of small inventories of segments in pronominal systems. Among consonants, *n*, *m*, *k* and *t* are highly favoured. Among vowels, *a* and *i* are highly favoured.

Nichols and Peterson (1996) use a larger sample. In their study the 1SG pronoun has *n* as the initial C in 37/173 languages (or 20.8%), 2SG has *n* as the initial in 23/173 languages (13.3%).

Given that singular pronouns are generally monosyllabic and that **n* and **k* are highly favoured consonants in pronoun systems it would seem we need to treat Greenberg's pronominal evidence for Indo-Pacific with some caution. Indeed there is other evidence showing that the chances of two languages independently developing 1SG, and 2SG pronouns beginning with the same consonant are by no means miniscule. It happens that Trans New Guinea 1SG, 2SG and 3SG independent pronouns have close matches in some languages of the Afro-Asiatic, Algonquian and Austronesian families. The following table compares the well attested pTNG forms with the independent pronouns of Hausa (Afro-Asiatic) and SW Ojibwe (Algonquian) and with preverbal subject pronouns in two Austronesian languages of Vanuatu: Mera Lava and Raga.

Table 4: Singular pronouns in languages of four unrelated families

	pTNG	Hausa	Ojibwe	Mera Lava	Raga
1SG	*na	ni	ni:n	na, no	na
2SG	*ga	kai	ki:n	ko	go
3SG	*ya	shi	wi:n	a	k-ea

The Hausa 1SG and 2SG forms continue Proto Chadic forms that are similar. The Proto Algonquian singular pronouns are: 1sg *ni:la, 2sg *ki:la, 3sg *wi:la, with P-Alg *l > n in Ojibwe, merging with reflexes of P-Alg *n (J. Blevins pers. comm.).

The Mera Lava and Raga subject pronouns continue Proto Austronesian (PAN) independent forms with a number of changes. In the first singular forms the initial *n* is not original. PAN *aku '1SG' became *au in Proto Oceanic (POc), with irregular loss of *-k. In

a number of Vanuatu and SE Solomons languages POC *au became *nau, possibly because the *n of a preceding transitive marker was reanalysed as part of the object pronoun (the same as the independent forms). The independent singular pronouns were then adopted as preverbal subject markers, with some phonological reduction ensuing. *nau reduced to *na* and *no* in the two languages in question, thus coming to closely resemble the pTNG form. The PAn 2SG root was a disyllable, *kaSu, which normally took a prefix *i that marked independent pronouns. The Mera Lava and Raga forms continue *kaSu regularly, with *S lost and *au becoming o, thus creating a monosyllabic form that closely resembled the pTNG form. In the third singular the PAn form *(si-)ia (yielding POC *ia) is quite similar to pTNG *ya, and this is continued in Mera Lava and Raga with some irregular developments. There are other cases of Austronesian languages in Indonesia that have independently developed three singular pronouns closely resembling those attributed to pTNG.

While these observations do not rule out the possibility that some or all of the pronominal resemblances between TNG and non-TNG groups are due to common origin they show that there is a reasonable chance that some or all of the resemblances, specially those between geographically well-separated groups, may also be due to non-genetic factors.

4. Timor-Alor and Halmahera agree in having a first person inclusive plural pronoun in *p. The Timor-Alor witness is a TNG language but there is no good reason to think that this is even an old TNG feature, so this resemblance is unlikely to be a shared inheritance.
5. Third person plural. *d* and *t* forms occur in:
 - New Britain: Taulil, Butam, Sulka *ta*.
 - Bougainville: Siwai *at*, Galeli *idu*. NNG: Sko *tea*, Sangke *te*, *ndu*, (*n*)*di*.
 Weak. Such resemblances between three widely separated groups are likely to be due to chance.
6. Suffixes on verb marking subject person and number. This is characteristic of most TNG groups. Greenberg notes that the structural type also occurs in Nimboran, of NNG, but such a structural resemblance in a language close to TNG languages is of little value. Any formal agreements between particular suffixes are confined to TNG.
7. Pronouns (i) prefixed to noun to indicate possessor, (ii) prefixed to verb to indicate object. These features are characteristic of the TNG family only. Given SOV order, it is not surprising that object pronouns precede the verb.
8. In three New Guinea groups (SWNG, SNG, CNG), all now assigned to the TNG family, certain tenses are marked by subject-tense portmanteau suffixes in which:
 - (i) second and third persons are identical in non-singular dual and plural
 - (ii) first person differs from non-first person by a vowel change which is the same for plural (and for dual if there is one). Recurrent variants are *a/i*, *e/i* and *i/e*. These two features, and especially *a/i* variation, may well be old TNG features but they are not attested in Indo-Pacific groups other than TNG.
9. A plural marker on nouns, *mana* or *mVnV*, occurs in some languages in three of Greenberg's groups: Timor-Alor (in Abui), Central New Guinea (Moni), and East New Guinea (in Binanderean). These are all TNG groups.

10. Marking of grammatical gender (or noun classes). Grammatical gender, done by vowel alternations, is a feature of a number of Indo-Pacific groups. In most groups the masculine vowel is more front than the feminine. Greenberg considers this correlation to be a major piece of evidence for his hypothesis. He discusses at some length gender marking in Marind, a TNG language of south central New Guinea. Marind has four genders: 1. masculine human, 2. feminine human and animals, 3. inanimate, 4. inanimate. The most basic pattern is: *e* masculine singular, *u* feminine singular, *a* inanimate class 1, *i* inanimate class 2 + plural of masculine and feminine. This pattern is manifested in some nouns such as *anem* 'man', *anum* 'woman', *anim* 'people' but more widely in adjectival agreement with nouns.

Within TNG gender-marking is virtually confined to the south central New Guinea area and there are no strong grounds for attributing it to pTNG. However, gender marking is found in several other Papuan groups. Halmahera uses consonant variation for this purpose. Gender marking is widespread in NNG languages, e.g., Monambo of the Sepik region, has a five gender system with feminine singular *u*, neuter singular *i*, and three consonantal markers. Taulil and Butam of New Britain have masculine *a*, feminine *e*, neuter *i* (a striking resemblance to Marind and Monambo) and plural *ta*.

In Bougainville Nasioi has contrasts like *nuring* 'son', *norang* 'daughter', *naung* 'husband', *naang* 'wife', where *i* marks masculine and *a* feminine.

The specific correlations that Greenberg points to are indeed striking and may be the shadowy remnants of an ancient shared history. However, without a cross-linguistic survey of the kind that has been done for pronoun forms it is hard to evaluate the chances of these sorts of resemblances arising independently. My impression is that vowel alternations are quite widely used to mark gender contrasts in determiners, pronouns and nouns.

11. Past tense marked by a suffix containing a velar consonant. This feature is found in some members of at least four TNG groups (SWNG, SNG, CNG, NENG) and such a distribution yields a fairly promising case for reconstructing pTNG *-k '(remote) past'. We also find *-ka* in most North Andaman languages and some Halmahera languages and forms containing *k* or *g* in some NNG languages and in Bilua, a Central Melanesian language. However, once again, in the absence of other, more convincing evidence for connecting AN, HA, NNG and CM it is difficult to place much weight on this resemblance. Given that suffixes tend to erode and that velar stops are not the most stable of consonants, the chances of any language retaining a past tense suffix based on a velar consonant for 30 or 40 millennia would seem to be very small.

To sum up, the grammatical evidence includes several morphological agreements that support a TNG group, namely items 1–3, 6–8, and 11, and perhaps 9. There is some shadowy evidence for connecting TNG with certain other New Guinea area groups and isolates. It would hardly be surprising if TNG shares a common ancestor with some other languages of the New Guinea mainland at a time depth of between 10 and 15 millennia, recent enough for a few traces to remain.

The evidence for relating either Tasmanian and Andaman to any of the other groups is negligible. The few resemblances are best viewed as accidental. The same assessment applies to resemblances between Bougainville and Central Melanesian and any of the other groups.

7 Why was Greenberg persuaded?

Finally, we are left with this question: Why was Greenberg, an extremely erudite and astute scholar, and the author of some highly regarded papers on the methodology of historical linguistics, persuaded that he had a pretty good case for Indo-Pacific when his critics are unimpressed by the evidence? Several factors can be readily discerned.

First, Greenberg seriously underestimated the chances of different languages independently developing resemblant pronoun forms and, more generally, resemblant lexical forms. Greenberg drew a very long bow in assigning putative cognates to his ‘Indo-Pacific etymologies’, allowing great latitude both in respect of formal and semantic variation. This weakness would no doubt have been corrected had he followed application of the multilateral method with statistical tests for significance. His reluctance to take this further step remains a puzzle.

Second, there is the subgrouping issue. Greenberg deserves credit for recognising a number of grammatical and lexical agreements that support what we now call the Trans New Guinea family. However, the evidence he compiled for uniting groups 2, 5, 8–10 and part of 6, i.e. the TNG languages, is much stronger than the evidence for any wider grouping among the 14 putative subgroups of Indo-Pacific. To the extent that there are resemblances among groups 1, 3, 4, 6, 11–14 and between these and the TNG groups, they are few in total and flimsy in quality. If Greenberg had counted the numbers of resemblances across his subgroups the differences would surely have been obvious but he did not provide any statistical arguments and it is possible that that he did not see the patterning. At any rate, with the benefit of hindsight we can see that his failure to identify the Trans New Guinea languages as a single primary unit in his subgrouping hypothesis, rather than as representing several coordinate subgroups, led him to overvalue the importance of agreements between the TNG groups as evidence for a wider Indo-Pacific stock.

Third, he did not try to support his etymologies by seeking recurrent sound correspondences, either within or between particular subgroups. Given the scope of Indo-Pacific, we can hardly blame Greenberg for not investigating sound correspondences — for most of the putative subgroups he could not have made much progress in such a task with the fragmentary data at his disposal and even with excellent data the job of analysing correspondences for all of the groups would be beyond any single person. However, it is not especially difficult to demonstrate recurrent sound correspondences between the better-known TNG languages. Greenberg himself could have done so for the languages which figure most prominently in his etymologies, had he chosen to undertake this step in the comparative method

Greenberg was critical of the categorisation of historical linguists into ‘lumpers’ versus ‘splitters’, arguing that the number of groups related under a hypothesis should not be an issue. But surely the central issue has always been the quality of the evidence. The difference is that lumpers are satisfied with a lesser standard of proof than splitters. It seems that, in the case of Indo-Pacific, Greenberg forgot his own wise advice, cited earlier in this paper, and which I repeat here:

... where more than one family is represented, ... the contrast between the relatively numerous and qualitatively superior resemblances between related languages, compared to the sporadic and qualitatively poorer resemblances among unrelated languages, becomes readily apparent. In this way the presence of unrelated languages provides a control for distinguishing mere chance from genetically significant resemblances. (Greenberg 2005:42)

I remain astonished that such a hugely experienced and perceptive scholar did not take a more cautious and critical view of the evidence before him. My hunch is that Greenberg's early successes in relating African groups made him addicted to the search for long range relationships and led him to take a less critical view of the evidence than he should have. Great scholars are not immune to hubris.

References

- Barker, Graeme, Tim Reynolds and David Gilbertson, eds. 2005. *The human use of caves in Peninsular and Island Southeast Asia*. Special Issue. *Asian Perspectives* 44(1).
- Bellwood, Peter. 1997. *The prehistory of the Indo-Malaysian archipelago*. (2nd edition.) Honolulu: University of Hawaii Press.
- Blevins, Juliette. 2007. A long lost sister of Proto-Austronesian? Proto-Ongan, mother of Jarawe and Ongan in the Andaman islands. *Oceanic Linguistics* 46(1).154–198.
- . To appear. Linguistic clues to Andamanese pre-history: understanding the North-South divide. In Tom Güldemann, Patrick McConvell and Richard Rhodes, eds *Hunter-gatherers and linguistic history: a global perspective*. Cambridge: Cambridge University Press.
- Campbell, Lyle. 1988. Review of *Language in the Americas* (Greenberg 1987). *Language* 64.591–615.
- . 1997. *American Indian languages: the historical linguistics of native America*. Oxford: Oxford University Press.
- Campbell, Lyle and William J. Poser. 2008. *Language classification, history and method*. Cambridge: Cambridge University Press.
- Capell, Arthur. 1948–49. Distribution of the languages in the Central Highlands, New Guinea. *Oceania* 19(2–4).114–129, 234–253, 349–377.
- Cavalli-Sforza, Luigi L. 2000. *Genes, peoples and languages*. San Francisco: North Point Press.
- Cosgrove, Richard, Jim Allen and Brendan Marshall. 1990. Palaeo-ecology and Pleistocene human occupation in south central Tasmania. *Antiquity* 64(242).59–78.
- Cowan, H.J.K. 1957. A large Papuan language phylum in West New Guinea. *Oceania* 28(2).159–166.
- . 1960. Nadere gegevens betreffende de verbreiding der West-Papoease taalgroep (Vogelkop, Nieuw-Guinea). *Bijdragen tot de Taal-, Land- en Volkenkunde* 116(3).350–364.
- Croft, William, ed. 2005. Editor's introduction. In Greenberg 2005, x–xxxvii.
- Crowley Terry and R.M.W. Dixon. 1981. Tasmanian. In R.M.W. Dixon and B. Blake, eds *Handbook of Australian languages*, vol. 2, 394–421. Canberra: The Australian National University Press.
- Dunn, Michael, Ger Reesink and Angela Terrill. 2002. The East Papuan languages: a preliminary typological appraisal. *Oceanic Linguistics* 41.28–62.
- Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley and Stephen C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309.2072–2075.

- Dyen I., A.T. James and J.W.L. Cole. 1967. Language divergence and estimated retention rate. *Language* 43(1).150–171.
- Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.
- 2000. The languages of New Guinea. *Annual Review of Anthropology* 357–404.
- 2005. Linguistic prehistory in the Sepik-Ramu basin. In Pawley et al., eds, 2005, 109–144.
- Franklin, Karl. 1973. Introduction. In K. Franklin, ed. *The linguistic situation in the Gulf District and adjacent areas, Papua New Guinea*. Canberra: Pacific Linguistics.
- Friedlaender, Jonathan, ed. 2007. *Population genetics, linguistics, and culture history in the southwest Pacific: a synthesis*. New York: Oxford University Press.
- Friedlaender, Jonathan S. Françoise R. Friedlaender, Floyd A. Reed, Kenneth K. Kidd, Judith R. Kidd, Geoffrey K. Chambers, Rodney A. Lea, Jun-Hun Loo, George Koki, Jason A. Hodgson, D. Andrew Merriwether, and James L. Weber. 2008. The genetic structure of Pacific Islanders. *PLoS Genetics* 4(1).e19.
- Gordon, Matthew. 1995. The phonological composition of personal pronouns: implications for genetic hypotheses. In Jocelyn Ahlers, L. Bilmes, J. Guenter, B. Kaiser and Ju Namkung, eds *Proceedings of the 21st annual meeting of the Berkeley Linguistic Society*, 117–128. Berkeley: Berkeley Linguistic Society.
- Green, Roger C. 2003. The Lapita horizon and traditions: signature for one set of Oceanic migrations. In C. Sand, ed. *Pacific archaeology: assessments and prospects*, 95–129. Les Cahiers de l'Archéologie en Nouvelle-Calédonie, vol. 15. Noumea: New Caledonia Museum.
- Greenberg, Joseph H. 1958. Report on the classification of the non-Austronesian languages of the Pacific. Mimeo. [Report to the Wenner–Gren Foundation for Anthropological Research.]
- 1960. Indo-Pacific etymologies. Mimeo.
- 1971. The Indo-Pacific hypothesis. In Thomas A. Sebeok, ed. *Current trends in linguistics 8: linguistics in Oceania*, 807–871. The Hague: Mouton.
- 1987. *Language in the Americas*. Stanford: Stanford University Press.
- 2005. *Genetic linguistics. Essays on theory and method*. (ed. William Croft.) Oxford: Oxford University Press.
- Groube, L.M., J. Chappell, J. Muke and D. Price. 1986. A 40,000 year old human occupation site at Huon Peninsula, Papua New Guinea. *Nature* 324.453–455.
- Haiman, John. 1979. Review of Wurm, ed. 1975. *Language* 55(4).894–903.
- Heeschen, Volker. 1978. Review of Wurm, ed. 1975. *Anthropos* 72.978–980.
- Kirch, Patrick V. 2000. *On the road of the winds: an archaeological history of the Pacific Islands before European contact*. Berkeley: University of California Press.
- Kruskal, B., I. Dyen and P.D. Black, 1971. The vocabulary method of reconstructing family trees: innovations and large scale applications. In F.R. Hodson, D.G. Kendall, and P. Tautu, eds *Mathematics in the archaeological and historical sciences*, 30–55. Edinburgh: Edinburgh Press.

- Lang, Ranier. 1976. Review of Wurm, ed. 1975. *Kivung*, *Journal of the Linguistic Society of Papua New Guinea* 9(1).72–80.
- Laycock, Donald C. 1975a. A hundred years of Papuan linguistic research: Eastern New Guinea. In Wurm, ed., 1975, 43–115.
- 1975b. The Torricelli Phylum. In Wurm, ed., 1975, 767–780.
- Laycock, Donald C. and John Z'graggen. 1975. The Sepik-Ramu Phylum. In Wurm, ed., 1975, 731–763.
- Leavesley, Matthew and John Chappell. 2004. Buang Merabak: additional early radiocarbon evidence for the colonisation of the Bismarck Archipelago, Papua New Guinea. *Antiquity* 78(301).
- Loukotka. 1957. Classification des langues Papoues. *Lingua Posnaniensis* (Poznan). 19–83.
- Manning, Patrick. 2006. *Homo sapiens* populates the Earth: a provisional synthesis, privileging linguistic data. *Journal of World History* 17(2).115–158.
- McElhanon, Kenneth A. 1967. Preliminary observations on Huon Pensinsula languages. *Oceanic Linguistics* 6.1–45.
- 1970. Lexicostatistics and the classification of the Huon Peninsula languages. *Oceania* 40(3).214–231.
- McElhanon, K. and C.L. Voorhoeve. 1970. *The Trans-New Guinea Phylum: explorations in deep-level genetic relationships*. Canberra: Pacific Linguistics.
- Mellars, Paul. 2006a. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *PNAS* 103(25).9381–9388.
- 2006b. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* 313, no. 5788.796–800.
- Mithun, Marianne. 1999. *The languages of Native North America*. Cambridge: Cambridge University Press.
- Mulvaney, John and Johan Kamminga. 1999. *Prehistory of Australia*. St Leonards, NSW: Allen and Unwin.
- Nichols, Johanna and David A. Peterson. 1996. The Amerind personal pronouns. *Language* 72.336–371.
- O'Connell, J.F. and J. Allen. 2004. Dating the colonization of Sahul (Pleistocene Australia-New Guinea): a review of recent research. *Journal of Archaeological Science* 31.835–853.
- O'Connor, Sue. 2007. New evidence from East Timor contributes to our understanding of earliest modern human colonisation east of the Sunda Shelf. *Antiquity* 81.523–535.
- Pagel, Mark. 2000. Maximum likelihood models for glottochronology and for reconstructing linguistic phylogenies. In Colin Renfrew, April MacMahon, Larry Trask, eds *Time-depth in historical linguistics*, 189–207. Cambridge: The McDonald Institute of Archaeology.
- Pagel, Mark, Quentin Atkinson and Andrew Meade. 2007. Frequency of word use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449.717–720.

- Pagel, Mark and Andrew Meade. 2006. Estimating rates of lexical replacement on phylogenetic trees of languages. In Peter Forster and Colin Renfrew, eds *Phylogenetic methods and the prehistory of languages*, 173–182. Cambridge: McDonald Institute.
- Pavlidis, Christina and Chris Gosden. 1994. 35000 year-old sites in the rainforests of West New Britain, Papua New Guinea. *Antiquity* 68.604–610.
- Pawley, Andrew. 1998. The Trans New Guinea Phylum hypothesis: a reassessment. In Jelle Miedema, Cecilia Ode and Rien A.C. Dam, eds *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, 655–689. Amsterdam: Editions Rodopi.
- 2001. The Proto Trans New Guinea obstruents: arguments from top-down reconstruction. In Pawley, Ross and Tryon, eds, 2001, 261–300.
- 2002. The Austronesian dispersal: languages, technologies, peoples. In Peter Bellwood and Colin Renfrew, eds *Examining the farming/language dispersal hypothesis*, 251–273. Cambridge: McDonald Institute of Archaeological Research, Cambridge University.
- 2005a. The chequered career of the Trans New Guinea hypothesis: recent research and its implications. In Pawley et al., eds, 2005, 67–107.
- 2005b. Papuan languages. In K. Brown and Anne Anderson, eds *Encyclopaedia of language and linguistics* (2nd edition), vol. 9, 162–171. Oxford: Elsevier.
- 2007. Recent research on the historical relationships of the Papuan languages, or, What does linguistics say about the prehistory of Melanesia? In Friedlaender, ed., 2007, 36–58.
- n.d. Some Trans New Guinea Phylum cognate sets. Unpublished Manuscript. Canberra: The Australian National University.
- Pawley, Andrew, Robert Attenborough, Jack Golson and Robin Hide, eds. 2005. *Papuan pasts: cultural, linguistic and biological histories of the Papuan-speaking peoples*. Canberra: Pacific Linguistics.
- Pawley, Andrew, Malcolm Ross, and Darrell Tryon, eds. 2001. *The boy from Bundaberg: studies in Melanesian linguistics in honour of Tom Dutton*. Canberra: Pacific Linguistics.
- Ray, Sidney H. 1919. A new linguistic family. *The Bible in the World*. October.
- Reesink G. 2005. West Papuan languages: roots and development. In Pawley et al., eds, 2005, 185–219.
- Renfrew, Colin, April McMahon and Larry Trask, eds. 1999. *Time depth in historical linguistics*. Cambridge: MacDonald Institute for Archaeological Research.
- Rhodes, Richard. 1997. On pronominal systems. In Iren Hegedüs, P. Michalove and A. Manster Ramer, eds *Indo-European, Nostratic and beyond, festschrift for Vitaliy V. Shevoroskin*, 293–319. J. Indo-European Studies monograph 22. Washington, D.C.: Institute for the Study of Man.
- Ringe, Don. 1992. *Calculating the factor of chance in language comparison*. Philadelphia: American Philosophical Society.
- 1996. The mathematics of 'Amerind'. *Diachronica* 13(1).135–154.
- 1999. How hard is it to match CVC-roots? *Transactions of the Philological Society* 97(2).213–244.

- Ross, Malcolm D. 1995. The great Papuan pronoun hunt: recalibrating our sights. In Connie Baak, Mary Bakker and Dick van der Meij, eds *Tales from a concave world: liber amicorum Bert Voorhoeve*, 139–168. Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University.
- 2000. Pronouns as preliminary evidence for grouping Papuan languages. Unpublished manuscript. Canberra: The Australian National University.
- 2001. Is there an East Papuan phylum? Evidence from pronouns. In Pawley, Ross and Tryon, eds, 2001, 301–321.
- 2005. Pronouns as preliminary evidence for grouping Papuan languages. In Pawley, et al., eds, 2005, 15–65.
- Ross, Malcolm and Åshild Næss. 2007. An Oceanic origin for Aiwoo, the language of the Reef Islands? *Oceanic Linguistics* 46(2).456–498.
- Ruhlen, M. 1991. *A guide to the world's languages*. vol. 1. *Language classification*. (2nd edition) Stanford: Stanford University Press.
- Schmidt, Wilhelm. 1952. *Die tasmanischen Sprachen*. Utrecht-Anvers: Spectrum.
- Specht, Jim. 2005. Revisiting the Bismarcks: some alternative views. In Pawley et al., eds, 2005, 235–288.
- Spriggs, Matthew. 1997. *The Island Melanesians*. Oxford: Blackwell.
- Thangaraj, K., G. Chaubey, A.G. Reddy, V.K. Singh and L. Singh. 2006. Unique origin of Andaman Islanders: insight from autosomal loci. *Journal Human Genetics* 51(9).800–804.
- Thangaraj, Kumarasamy, Lalji Singh, Alla G. Reddy, V.Raghavendra Rao, Subhash C. Sehgal, Peter A. Underhill, Melanie Pierson, Ian G. Frame and Erika Hagelberg. 2003. Genetic affinities of the Andaman Islanders, a vanishing human population. *Current Biology* 13(2).86–93.
- Torrence, Robin, Vince Neall, Trudy Doelman, Ed Rhodes, Chris McKee, Hugh Davis, Roberto Bonetti, Alessandra Guglielmetti, Alberto Manzoni, Massimo Oddone, Jeff Parr and Wallace Cleland. 2004. Pleistocene colonisation of the Bismarck Archipelago: new evidence from New Britain. *Archaeology in Oceania* 39(3).101–130.
- Viberg, Åke. 1984. The verbs of perception: a typological study. In Brian Butterworth, Bernard Comrie and Östen Dahl, eds *Explanations for language universals*, 123–262. Berlin: Mouton de Gruyter.
- Voegelin, Carl F. and Florence M. Voegelin. 1965. Languages of the World. Indo-Pacific fascicle 5. *Anthropological Linguistics* 7(9).1–114.
- Voorhoeve, C.L. 1968. The Central and South New Guinea Phylum: a report on the language situation in south New Guinea. In *Papers in New Guinea Linguistics* 8, 1–17. Canberra: Pacific Linguistics.
- 1975. Central and western Trans-New Guinea Phylum languages. In Wurm, ed., 1975, 345–459.
- 2005. Asmat-Kamoro, Awyu-Dumut and Ok: an enquiry into their linguistic relationships. In Pawley et al., eds, 2005, 45–166.
- Wickler, S. and M. Spriggs. 1988. Pleistocene occupation of the Solomon Islands, Melanesia. *Antiquity* 62.703–706.

- Whitehouse, Paul, T. Usher, M. Ruhlen and W.S-Y Wang. 2004. Kusunda: an Indo-Pacific language in Nepal. *PNAS* 101(15).5692–5.
- Wurm, Stephen A. 1954. Notes on the structural affinities of non-Melanesian (Papuan) languages. In W. Koppers, ed. *Die Wiener Schule der Voelkerkunde, Festschrift zum 25. jährigen Bestand 1929-1954*, 467–472. Vienna: Herold.
- 1960. The changing linguistic picture of New Guinea. *Oceania* 31(2).121–136.
- 1964. Australian New Guinea Highland languages and the distribution of their typological features. *American Anthropologist* 66(4-part 2).79–97.
- 1965. Recent comparative and typological studies in Papuan languages of Australian New Guinea. *Lingua* 15.373–399.
- 1971. The Papuan linguistic situation. In T.E. Sebeok, ed. *Current trends in linguistics* vol. 8. *Oceania*, 541–657. The Hague: Mouton.
- 1975a. Possible wider connections of Papuan languages: Papuan and Australian; Greenberg's hypothesis. In Wurm, ed., 1975, 925–932.
- 1975b. Personal pronouns. In Wurm, ed., 1975, 191–217.
- 1982. *Papuan languages of Oceania*. *Ars Linguistica* 7. Tübingen: Gunter Narr.
- Wurm, Stephen A., C.L. Voorhoeve and K. McElhanon. 1975. The Trans New Guinea Phylum in general. In Wurm, ed., 1975, 299–322.
- Wurm, Stephen A., ed. 1975. *New Guinea area languages*. vol. 1. *Papuan languages and the New Guinea linguistic scene*. Canberra: Pacific Linguistics.
- Wurm, Stephen A. and Shiro Hattori, 1981–83. *Language atlas of the Pacific area* (2 vols). Canberra: Australian Academy for the Humanities in collaboration with the Japanese Academy.
- Z'graggen, John. 1975. The Madang-Adelbert Range sub-phylum. In Wurm, ed., 1975, 569–612.